ABSTRACT

DEPARTMENT OF COMPUTER AND INFORMATION SCIENCE

HEWA BOSTHANTHIRIGE, MIHIRI B.Sc. UNIVERSITY OF KELANIYA, 2006

M.S. UNIVERSITY OF COLOMBO, 2011

M.S. CLARK ATLANTA UNIVERSITY, 2013

OUTLIER DETECTION IN NETWORK DATA USING THE BETWEENNESS CENTRALITY

Committee Chair: Roy George, Ph.D.

Thesis dated July 2015

Outlier detection has been used to detect and, where appropriate, remove anomalous observations from data. It has important applications in the field of fraud detection, network robustness analysis, and intrusion detection. In this paper, we propose a Betweenness Centrality (BEC) as novel to determine the outlier in network analyses. The BEC of a vertex on a graph is a measure for the participation of the vertex in the shortest paths in the graph. The BEC is widely used in network analyses. Especially in a social network, the recursive computation of the BEC of vertices is performed for the community detection and finding the influential user in the network. In this paper, we propose that this method is efficient for finding outlier in social network analyses. Furthermore we show the effectiveness of the new methods.

OUTLIER DETECTION IN NETWORK DATA USING THE BETWEENNESS CENTRALITY

A THESIS

SUBMITTED TO THE FACULTY OF CLARK ATLANTA UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE MASTER OF SCIENCE

BY

MIHIRI HEWA BOSTHANTHIRIGE

DEPARTMENT OF COMPUTER AND INFORMATION SCIENCE

ATLANTA, GEORGIA

JULY 2015

ACKNOWLEDGEMENTS

The work presented here would not have been possible without the support and guidance of several people. First of all, I would like to express my greatest appreciation to my advisor, Dr. Roy George, for his invaluable guidance, patience, motivation, enthusiasm, immense knowledge, and supervision throughout the tenure of my study. He is a great teacher who is sharing his knowledge with students. I would also like to thank committee members, Dr. Khalil Shujaee and Dr. Hsin-Chu Chen, for their advice and suggestions. I acknowledge the support given by the faculty and staff of the Department of Computer and Information Science, and especially Research Scientists Mr. Ali Sazegarnejad. This research was partially supported by funds from the Army Research Laboratory Grant No: W911NF-12-2-0067 and the Army Research Office Contract Number W911NF-11-1-0168.

Last but not the least, I would like to thank my family; my husband Asanga; my parents; brother; sister; sister-in-law; and my friends for spiritually supporting me throughout my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BEC	Betweenness Centrality
SNN	Shared Nearest Neighbor
m-SNN	Modified Shared Nearest Neighbor
SVM	Support Vector Machines
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative

CHAPTER 1

INTRODUCTION

Outlier detection is an important data mining task that is focused on the discovery of objects that are exceptional when compared with a set of observations that are considered typical. In many data analysis tasks, a large number of variables are being recorded or sampled. One of the first steps towards obtaining a coherent analysis is the detection of outlaying observations. Although outliers are often considered an error or noise, they may carry important information. Detected outliers are candidates for aberrant data that may otherwise adversely lead to model misspecification, biased parameter estimation, and incorrect results. These objects are important since they often lead to the discovery of exceptional events. Substantial research has been done in outlier detection and these are classified into different types with respect to the detection approach being used. Exemplar techniques include classification-based methods, nearest neighbor-based methods, cluster-based methods and statistical-based methods (19). In the classification-based approach (31, 32) a model is created from a set of labeled data points and then a test point is classified into one of the classes using appropriate testing. Support Vector Machine (SVM) based methods (30), methods based on neural networks (33) and Bayesian networks-based methods (25, 28, 34) belong to classification based technique. The testing phase of this method is considerably fast as each test data is compared against

the prebuilt model. The accuracy of classification-based methods rely on the availability of accurate preclassified examples for different normal classes, which is rarely found. Nearest neighbor-based methods (27, 29, 35) involve distance or similarity measures which is defined between data points. In this thesis, we discuss a new method to find an outlier that is based on a graph. This method efficiently reduces the search space by finding a candidate set of vertices whose betweenness centralities can be computed using candidate vertices only.

The Betweenness Centrality (BEC) is a measure that computes the relative mportance of a vertex in a graph, and it is widely used in network analyses such as social network analysis, biological graph analysis, and road network analysis (1). In the social network analysis, a vertex with higher centrality can be viewed as a more important vertex than a vertex with lower centrality. The BEC of a vertex in a graph is a measure used for the participation of the vertex in the shortest paths in the graph. There are many previous works on the BEC problem. The concept of the BEC is proposed in (35), but the definition proposed in (40) is more widely used. Recently, many variants of the definition are proposed in (38), (37) improves the computation time of the BEC based on a modified breadth-first search algorithm and the dependency of a vertex, and it is the fastest known algorithm that computes the exact BEC of all the vertices in a graph. The computations of the shortest paths between all pairs of vertices are time consuming. Therefore, another definition of BEC is proposed (22); this based on a random walk. In (42) each vertex has a probability of visiting its neighbor vertices. Also, (36, 39, 41) propose approximation

algorithms for computing the betweenness centrality (43) and (44) adopt the betweenness centrality for detecting communities in a social network.

Although many methods currently exist on calculating the BEC and the BEC is one of the major methods used in analyzing social network graphs, none of the existing methods address the problem of updating BEC. In this thesis we propose the betweenness centrality to find out outliers for network type data.

The next section of this thesis describes related terms and definitions which are used throughout the thesis. Furthermore, it outlines the approach that explains the algorithm behind the BEC approach. To get a better understanding and to demonstrate the accuracy of BEC, several experiments were conducted with different kinds of synthetic data sets which are described in detail in the experimental results section. We apply the BEC technique to find outliers in synthetic data sets and compare it with another an alternate technique the modified-Shared Nearest Neighbor(m-SNN) (3). Finally, we conclude the thesis with a discussion of the performance, accuracy, and the importance of the proposed technique. From the results of experiments, it is clear that this technique gives better results in comparison to the m-SNN by giving higher true positive and true negative values and very low false positive and false negative values for network type data.

The m-SNN (modified-Shared Nearest Neighbor) method (3) is based on the nonparametric clustering algorithm, the Shared Nearest Neighbor (SNN) approach developed by Ertöz et al. (9). This method, we consider the ratio between the summation of Euclidean distances to shared nearest neighbors and their total number of shared

neighbors. To differentiate between outliers and normal nodes, hypothesis testing is used, Babara et al (18) and Rogers (4).

The outline of the thesis is as follows. Chapter 1 of this work introduces the topic and provides background and related work on outlier detection. Chapter 2 describes related terms and definitions which are used throughout the thesis. Chapter 3 outlines the approach that explains the algorithm behind BEC approach. To get a better understanding and to demonstrate the accuracy of BEC, several experiments were conducted with different kinds of synthetic data sets those are described in more detail in Chapter 4. Chapter 5 concludes the research with a discussion of the performance, accuracy and the importance of the proposed technique. From the results of experiments, it is clear that this technique gives better results in comparison to m-SNN by giving higher true positive and true negative values and very low false positive and false negative values.

CHAPTER 2

TERMS AND DEFINITIONS

We define Betweenness Centrality, Adjacency Matrices, and we define the terms p-value, null hypothesis (H_0) and alternative hypothesis (H_a) relating to the proposed technique.

Definition 1: Betweenness Centrality– A measure that computes the relative importance of a vertex in a graph. The formal definition is follows.

A graph is represented by $G=(V,E)$, where V is the set of vertices, and $E \subseteq V \times V$ is the set of edges. A path in a graph is represented by a sequence of vertices, $(v_1,...,v_n)$ where $v_i, v_j \in V$ for $1 \leq i, j \leq n$, $i \neq j$, except possible $1=n$. The betweenness centrality of a vertex $v_j \in G$ is:

$$c(v_j)=\sum_{i.k} \frac{\sigma_{v_i,v_k}(v_j)}{\sigma_{v_i,v_k}(1)}$$

Where v_i, v_j, v_k, $\in V$, $i \neq j \neq k$, $\sigma_{v_i,v_k}(v_j)$ is the number of shortest paths between v_i and v_k that include v_j, and σ_{v_i,v_k} is the number of shortest paths between v_i and v_k. The betweenness centrality can be computed as follows:

1. For each pair of vertices (v_s and v_t), compute the shortest paths between the two vertices.

2. For each pair of vertices, compute the ratio of each vertex participating in the shortest path(s). The ratio is the number of shortest paths between v_s and v_t that go through v_j divided by the number of shortest paths between v_s and v_t.

3. Accumulate the ratio for all pairs of vertices.

Definition 2: Adjacency Matrices – The adjacency matrix of a finite graph *G* on n vertices is the $n \times n$ matrix where the non-diagonal entry a_{ij} is the number of edges from vertex *i* to vertex *j*, and the diagonal entry a_{ii}, depending on the convention, is either once or twice the number of edges (loops) from vertex *i* to itself. Undirected graphs often use the latter convention of counting loops twice, whereas directed graphs typically use the former convention.

Figure 2 shows the adjacency matrix for undirected graph. A, B, C, D, E, and F represent the nodes. In the diagonal, all values are zero and if two nodes are connected; the matrix is denoted by the value of 1.

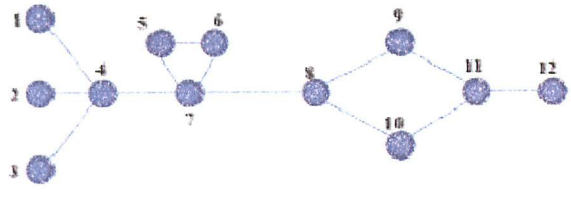

Figure 1: Shortest paths through nodes in destination IP.

	A	B	C	D	E	F
A	0	1	1	1	1	-
B	1	0	-	1	1	-
C	1	-	0	-	-	1
D	1	1	-	0	-	1
E	1	1	-	-	0	1
F	-	-	1	1	1	0

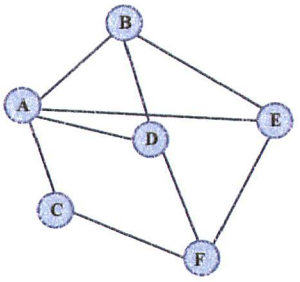

Figure 2: Undirected graph with adjacency matrix.

Definition 5: Null and Alternate Hypothesis

The null hypothesis and alternative hypothesis statements for BEC method are expressed below.

- Null Hypothesis = H_0: Point u is not an outlier (p-value >= τ)
- Alternative Hypothesis = H_a: Point u is an outlier (p-value < τ)

The p-value is the maximum probability of observing a test statistic as the null hypothesis is true. p-value is also known as observed level of significance while τ is the actual significance level. In (14), the p-value is obtained as the fraction of points in the class that have strangeness greater than or equal to that of the point. According to new algorithem algorithm, p-value of a point is calculated as the fraction of points in the class that have sparseness less than to that of the corresponding point. Therefore a larger p-value indicates the high probability of accepting null hypothesis, where as smaller p-value implies the high probability of rejecting null hypothesis, and accepting the alternative hypothesis.

CHAPTER 3

APPROACH

This outlier detection method is based on BEC for network data and p-value technique of hypothesis testing for finding outliers. Furthermore BEC methods can be directly finding outliers without using p-value technique. In results section it explain clearly. For each data point, we calculate its BEC by using adjacency matrix for network data. To find out the adjacency matrix for the data set, we calculate the shortest paths through nodes in the destination IP. Figure 1 show the shortest path through nodes in destination IP address. The numbers represent the label of each node for the given data points. The shortest path that is calculated creates an adjacency matrix from it by utilizing sparse matrices in order to increase computational speed. Our calculation is based on undirected network type data. The calculation for adjacency matrix yields an adjacency matrix from friendship nominations stored as a sparse matrix. The resulting adjacency matrix will include Id numbers in the first row and first column; it is shown in figure 3. To find the BEC, the calculation of the influence domain of each node in a given adjacency matrix for a given step, returns the undirected BEC for each node of undirected adjacency matrix 'adj'. Matrix 'adj' must be an undirected network and may or may not be sparse. The matrix is simple to change if the graph is directed.'adj' is assumed to have id numbers in

the first row and also, this code could probably be more vectorised to speed up calculations for large adjacency matrices.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & 0 & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & 0 & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & 0 & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & 0 \end{pmatrix} \end{array}$$

ID number	Adjacency matrix value
(1,1)	a_{11}
(1,2)	a_{12}
(1,3)	a_{13}
(1,4)	a_{14}
(1,5)	a_{15}
....	...
(5,4)	a_{54}
(5,5)	a_{55}

Figure 3: The resulting adjacency matrix including Id numbers in the first row and first column.

As our method needs to find the adjacency matrix for each data point, it is required to calculate the shortest path between each other data points.

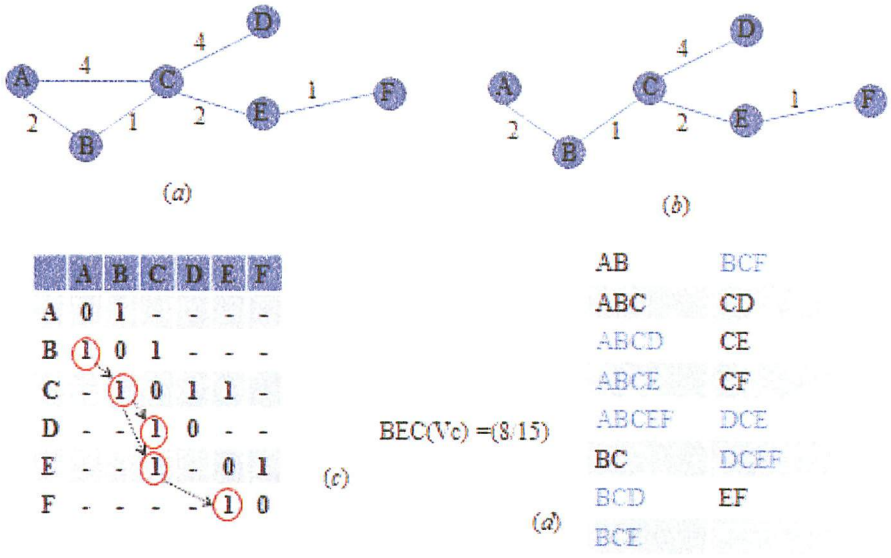

Figure 4: Betwenness centrality for node C step by step. (a) calculated Euclidean distance, (b) Shortest distance with Floyd algorithm, (c) Adjacency matrix.

In figure 4 shows six vertex labeling A, B, C, D, E, and F, by calculating betweenness centrality for vertex C. Figure 4(a) shows calculated Euclidean distance between nodes and using Floyd algorithm compute the shortest path for this graph (Figure 4(b)). According to the shortest path in the graph compute the adjacency matrix which is shown in figure4(c).In the adjacency matrix on the above, if the ties did not have a weight assigned to them, The connections are represent with 0s and 1s. In figure4 (d), the blue colored paths represent the 8 shortest paths in the graph that pass through C nodes. This will give this node a betweenness score of 8/15.

Figure 5: Betwenness centrality for node C step by step in Adjacency matrix.

In figure 4(c) show adjacency matrix for the sample graph and figure 5 shows how to compute the number of shortest path that include node C. To calculate number of shortest path including node C, it start from 1st node which is node A. There is no any shortest path starting with C node.

Since we have n data points, the complexity of calculating the shortest path is $O(n^2)$. Finally to find outliers we need to compare each data point with the other data points, thus resulting in $O(n^2)$ complexity.

Table 1: Betweenness Centrality Based Outlier Detection Algorithm

Procedure: Betweenness centrality Based Outlier Detection
Inputs: data[], a set of network data points;
Output: List of Outliers

// Finding Adjecency matrix for all the data points

Inputs: data[], Adjacency matrix for data points;
Output: List of Betweenness centrality for all data points

// Finding Betweenness centrality for all the data points

Inputs: data[], Betweenness Centrality for data points;
Output: List of Outliers
//Finding the outliers based on p-value method

CHAPTER 4

EXPERIMENTING WITH SYNTHETIC DATASETS

This section describes experiments and results with synthetic data sets followed by how the data was generated. We ran the experiments where τ was taken as 0.05. i.e., these experimental results are with 95% confidence.

4.1 Synthetic Data generation

To cover the broad range of applications, network type data sets were generated. We apply a rigorous set of tests to the data in the path to understand the strength or weakness of the method. In all cases we use probabilistic distribution based data generation which takes user inputs to decide parameters of the data pattern. i.e., identify variables and then use a probabilistic model to generate the required number of data points and outliers.

After generating data, each set of data points with scaling features were tested by using both the BEC method and m-SNN (3) outlier detection method. The m-SNN method is a modification of the SNN (Shared Nearest Neighbor) method that aids in outlier detection.

In this analysis, we generated network data sets of three different sizes viz. small (100<), medium (100< medium< 1000) and large (1000>). An example for a small data

set is a set with 56 total data points, where 6 of them were generated as global outliers which is small data set. After applying our new BEC method and m-SNN method with τ 0.05, all the expected global outliers were detected for the BEC method. Though the m-SNN approach was able to detect all the above labeled outliers correctly too, the results were not as accurate or precise as the BEC method

Table 2: Experimental results- BEC for synthesis data 21 points. Red dash circles show the outliers.

Node No:	Source IP	Destination IP	BEC
1	224.0.2.24	1.0.12.5	0
2	224.0.2.24	1.0.20.130	0.0224
3	224.0.2.24	1.0.0.8	0
4	224.0.2.24	1.0.0.6	0
5	224.0.2.24	1.0.0.131	4.9154
6	224.0.2.24	1.0.0.130	0.0001
7	224.0.2.24	1.0.0.133	0
8	224.0.2.24	1.0.0.5	0.0001
9	224.0.2.24	1.0.21.130	4
10	224.0.2.24	1.0.2.131	0.0224
11	224.0.2.24	1.0.20.9	0
12	224.0.2.24	1.0.0.9	0
13	224.0.2.24	1.0.0.134	4.9154
14	224.0.2.24	1.0.0.7	0.0001
15	224.0.2.24	1.0.0.132	0
16	224.0.2.24	1.0.2.6	0
17	224.0.2.24	1.0.12.132	4
18	224.0.2.24	1.0.2.5	0.0224
19	224.0.2.24	1.0.22.5	0
20	224.0.2.24	1.0.3.8	0
21	224.0.2.24	1.0.10.9	4.9154

The results obtained are summarized in Table 2 to demonstrate the calculated BEC for synthesis data point 21 and it show the Source and destination IP associate with done number. It clearly shows the relationship between BEC and outliers. BEC is always between 0 and 1 if any points not in that range are an outlier. Figure 4 shows the graph of calculated Betweenness centrality vs Node no according to Table 2 and this graph clearly shows to those five outliers in the data set.

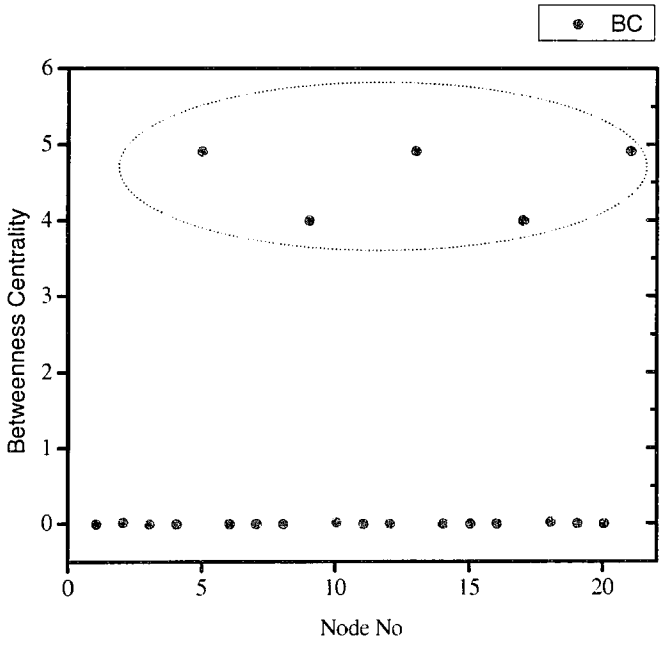

Figure 6: Outlier detection: BEC Technique – Synthetic data with 21 data points. Red dash circle show the outliers of the data set.

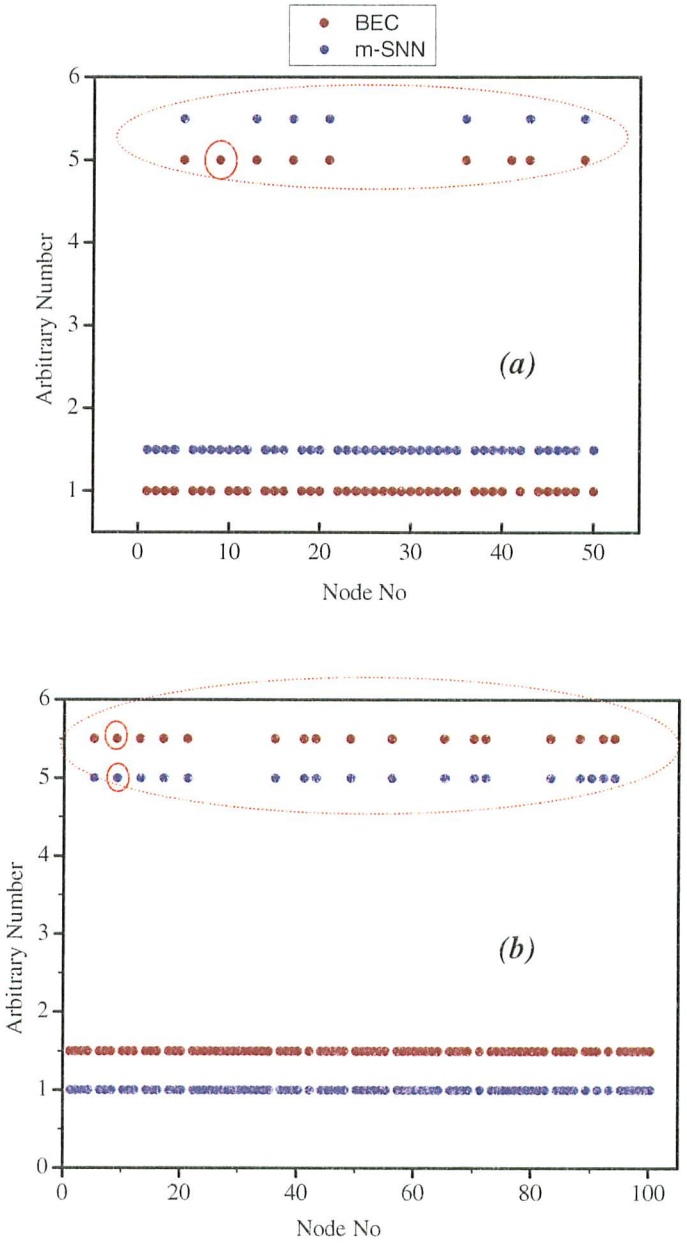

Figure 7: Outlier detection: BEC Technique and n-SNN– Synthetic data (a) with 50 and (b) 100 data points. Blue and Brown color shows results data for BEC and m-SNN respectively.

The results obtained are summarized in Figure 5 to demonstrate the comparison between BEC method and m-KNN for synthesis data point 50, and 100 and it show the

outliers in red circles. To compare these two techniques, assign arbitrary values for node numbers. We give arbitrary vales as 1 and 1.5 used for non-outliers, 5 and 5.5 used for outliers for BEC and m-SNN respectively. In Figure 5(a) have 50 data points for test run and its gives 9 and 7 data points outliers for BEC and m-SNN respectively. According to the results node no 9 is not an outlier for 50 data points' on the other hand with100 points test run, node number 9 is an outlier for m-SNN. However in BEC, number of number of data point is not depending on the outliers. According to graphs clearly show BEC has high accuracy compare with m-SNN.

The results obtained are summarized in Table 3 to demonstrate True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) values as percentages. It shows the average results for three different sizes of data sets. From the results, it is clear that the BEC has very high TP, TN percentages and very low FP, FN percentages compared to the m-SNN approach. Also the proposed method has the best results for the network type data. On comparing the results of complex path data sets, it is evident that the BEC is more robust in finding outliers (compared to m-SNN) particularly with respect to true positives and minimizing false negatives.

Table 3: Experimental results for BEC and m-SNN.

Technique	TP(%)	FP(%)	TN(%)	FN(%)
BEC	100.0	0.5	99.2	0.2
m-SNN	100.0	3.5	96.5	2.3

According to the tabulated results of Table 3, it is clear that BEC has very high TP, TN percentages and very low FP, FN percentages. Therefore BEC is enhanced in accuracy and performance when detecting outliers comparing to m-SNN approach.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

In this thesis, we have described an algorithm based on graph theory capable of detecting outliers in different types of network type data sets. This method is a combination of adjacency matrix and betweeness centralities which avoids assumptions about data distributions and uses hypothesis testing to detect outliers. Through a series of experiments, we have shown that this method achieves good results with very high true positive and true negative values with the BEC approach producing outlier detection results equivalent or better than the m-SNN method. Currently we are reformulating the algorithm to improve the run time efficiencies and also to parallelize the code to make it amenable for massive data sets. Furthermore, modifying this method can be used to identify an outlier to update a social network graph.

Currently we are reformulating the algorithm to improve the run time efficiencies and also to parallelize the code to make it amenable to massively large data sets. Also to reduce the time complexity of the BEC algorithm, our next step is to develop this method to find outlier without using p-value technique. Moreover we plan to continue our experimentations with more real datasets.

REFERENCES

1. M. J. Lee, R. H. Choi, J.Lee, C. W. Chung and J. Y. Park, "QUBE: a Quick algorithm for Updating BEtweenness centrality," World Wide Web (WWW). ACM, pp. 351-360, 2012.
2. D. Barbará, C. Domeniconi and J. P. Rogers, "Detecting outliers using transduction and statistical testing," ACM SIGKDD international conference on Knowledge discovery and data mining (KDD), pp. 55-64, 2006.
3. K. P. Liyanage, R. George and K. Shujaee, "Outlier Detection in Spatial Data using the m-SNN Algorithm," IEEE southeastCon 2013.
4. J. P. Rogers "Detection of Outliers in Spatial-temporal Data", PhD Thesis.
5. D. Hawkins, "Identification of Outliers," Chapman and Hall, London, 1980.
6. T. Johnson, I. Kwok and R. Ng, "Fast Computation of 2- Dimensional Depth Contours," Proc. 4th Int. Conf. on Knowledge Discovery and Data Mining, New York, NY, AAAI Press, 1998, pp. 224-228.
7. E. M. Knorr and R. T. Ng, "Algorithms for Mining Distance- Based Outliers in Large Datasets," Proc. 24th Int. Conf. on Very Large Data Bases, New York, NY, 1998, pp. 293-298.

8. E. M. Knorr and R. T. Ng, "Finding Intensional Knowledge of Distance-based Outliers," Proc. 25th Int. Conf. on Very Large Data Bases, Edinburgh, Scotland, 1999, pp. 211-222.

9. E. Levent, M. Steinbach and V. Kumar, "A New Shared Nearest Neighbor Clustering Algorithm and its Applications."

10. W. Wang, J. Yang and R. Muntz, "STING: A Statistical Information Grid Approach to Spatial Data Mining," Proc. 23th Int. Conf. on Very Large Data Bases, Athens, Greece, Morgan Kaufmann Publishers, San Francisco, CA, 1997, pp. 186-195.

11. T. Zhang , R. Ramakrishnan and M. Linvy, "BIRCH: An Efficient Data Clustering Method for Very Large Databases," Proc. ACM SIGMOD Int. Conf. on Management of Data, ACM Press, New York, 1996, pp.103-114.

12. F. Angiulli and C. Pizzuti, (2005) Outlier mining in large high-dimensional data sets. IEEE Transactions on Knowledge and Data Engineering, 17(2): 203-215.

13. A. Gammerman and Vovk, V. (2002) Prediction algorithms and confidence measures based on algorithmic randomness theory. Theoretical Computer Science. 287: 209-217.

14. UCI Machine Learning Repository http://www.ics.uci.edu/mlearn/MLRepository.html

15. V. Vapnik, (1998). Statistical Learning Theory. New York: Wiley.

16. M. Breunig, H. Kriegel, R. Ng and J. Sander, (2000) LOF: Identifying Density-Based Local Outliers. Proc. of the ACM SIGMOD Conference on Management of Data, 427- 438.

17. K. Proedru, I. Nouretdinov, V. Vovk and A. Gammerman, (2002) Transductive confidence machine for pattern recognition. Proc. 13th European conference on Machine Learning. 2430:381-390.

18. D. Barbara, C. Domeniconi and J. P. Rogers, "Detecting Outliers using Transduction and Statistical Testing," KDD'06, Philadelphia, Pennsylvania, 2006.

19. D. Velegrakis, "Outlier Detection over Data Streams using Statistical Modeling and Density Neighborhoods," Masters Thesis.

20. Eleazar Eskin. Anomaly detection over noisy data using learned probability distributions. Pages 255-262. Morgan Kaufmann, 2000.

21. E. Eskin, W. Lee, and S. J. Stolfo, " Modeling system calls for intrusion detection with dynamic window sizes," In proceedings of DARPA information Survivability Conference and Exposition 2 (DISCEX, 2001).

22. M. Ester, H-P Kriegel, J. Sander and X. Xu, "A density based algorithm for discovering clusters in large spatial databases with noise," In Proc. Of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96), pages 226-231, 1996.

23. L. Ertoz, M. Steinbach and V. Kumar, "Finding topics in collections of documents: A shared nearest neighbor approach," In Workshop on Text Mining, held in

conjunction with the First SIAM International Conference on Data Mining (SDM 2001). Society for Industrial and Applied Mathematics, 2003.

24. S. Guha, R. Rastogi and K. Rock, "A robust clustering algorithm for categorical attributes," Inf. Syst., 25(5): 345-366, 2000.

25. D. Barbara, N. Wu, and S. Jajodia, "Detecting novel network intrusions using bayes estimators," In Proceedings of the First SIAM Conference on Data Mining, April 2001.

26. R. Bronstein, J. Das, M. Duro, R. Friedrich, G. Kleyner, M. Muller, S. Singhal and I. Cohen, "Self-aware services: Using Bayesian networks for detecting anomalies in internet-based services," In Northwestern University and Stanford University Gary Igor, Pages 623-638, 2001.

27. E. M. Knorr, R. T. Ng, and V. Tucakov, "Distance-based outliers: algorithms and applications," The VLDB Journal, 8(3-4): 237-253, 2000.

28. M. Markou and S. Singh, "Novelty detection," A review – part 1: Statistical approaches. Signal Processing, 83: 2003, 2003.

29. J. C. Maxwell, "A Treatise on Electricity and Magnetism," 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

30. I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

31. K. Elissa, "Title of paper if known," unpublished.

32. R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

33. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

34. M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

35. J. Anthonisse and S. M. C. A. A. M. besliskunde, "Therush in a directed graph," Technical report, 1971.

36. D. A. Bader, S. Kintali, K. Madduri, and M. Mihail, "Approximating betweenness centrality," In Proceedings of the 5th international conference on Algorithms and models for the web-graph, WAW'07, pages 124–137, Berlin, Heidelberg, 2007. Springer-Verlag.

37. U. Brandes, "A faster algorithm for betweenness centrality." Journal of Mathematical Sociology, 25(1994):163–177, 2001.

38. U. Brandes, "On variants of shortest-path betweenness centrality and their generic computation," Social Networks, 30(2):136–145, 2008.

39. U. Brandes and C. Pich, "Centrality estimation in large networks" International Journal Of Bifurcation And Chaos, 17(7):2303, 2007.

40. L. C. Freeman, "A set of measures of centrality based on betweenness," Sociometry, 40(1):35–41, 1977.

41. R. Geisberger, P. Sanders, and D. Schultes, "Better approximation of betweenness centrality," In J. I. Munro and D. Wagner, editors, ALENEX, pages 90–100. SIAM, 2008.

42. M. E. J. Newman. "A measure of betweenness centrality based on random walks," Social Networks, 27(1):39–54, 2005.

43. M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," Physical Review E, 69(2): 26113, 2004.

44. J. W. Pinney and D. R. Westhead, "Betweenness-based decomposition methods for social and biological networks," In Interdisciplinary Statistics and Bioinformatics, pages 87–90. Leeds University Press, 2006.

LIST OF TABLES